**Step 10 — Γ-Robust VRPTW (Budget of Uncertainty)**

**Abstract (≈120 words)**

We introduced a Γ-robust routing layer that protects each route against a bounded number of worst-case travel-time shocks. Using a hybrid construction—per-arc quantile inflation (q=1.645 ≈ 95% one-sided) plus a route-level uncertainty budget Γ—we solved all 56 Solomon instances with Γ∈{1,2}. Compared to deterministic (DET) and quantile-buffer (Q120), Γ-robust plans are more conservative: they typically add vehicles and distance, and on the most time-tight random sets (R1, RC1) feasibility can drop at higher Γ. On clustered families (C1/C2) and long-horizon sets (R2/RC2), Γ-robust remained feasible while providing the strongest “worst-case” service guarantees. We then evaluated all methods under 200 correlated scenarios and added Γ runs to the common comparison tables. Γ-robust is the right choice when SLA compliance under spikes is paramount and extra fleet/cost is acceptable.

**What we implemented**

* **Method:** Γ-robust VRPTW, inspired by budget-of-uncertainty (Bertsimas–Sim).
* **Hybrid construction:**
  1. **Quantile inflation (q)** per arc to cover typical high traffic (one-sided 95%: q=1.645).
  2. **Budget Γ per route:** assume at most Γ legs simultaneously hit their worst deviations; enforce feasibility under those worst Γ arcs.
* **Solver:** OR-Tools (RoutingModel + time windows + capacity) with your existing metaheuristic flag (--meta GLS) and cost model distance + 10000\*vehicles.
* **Uncertainty link to Step 7:** We keep the same variability scale you adopted for scenario evaluation (cv\_global=0.20, cv\_link=0.10) so the robust buffers are consistent with your simulated test conditions.

**Key parameters used**

* --gamma {1,2} (how many legs per route may go worst-case)
* --mode hybrid (combines per-arc q-inflation with Γ budget checks)
* --q 1.645 (≈95% one-sided quantile)
* --cv\_global 0.20, --cv\_link 0.10 (calibration of arc deviations)
* --time\_limit 30, --vehicle\_cost 10000, --meta GLS

**Why we did it**

Deterministic and simple quantile buffers protect *on-average* or *high-percentile* traffic, but they don’t explicitly cap *how many* severe shocks may hit the same route. Γ-robustness says: “Even if up to Γ arcs on this route go bad at once, the plan still meets windows.” This is the natural choice for **SLA-driven operations** (e.g., “still on time if two segments get jammed”). It gives you a **single knob (Γ)** to trade cost/vehicles for resilience.

**What we ran (commands)**

You executed two robust configurations across all 56 instances and added them to the common evaluator:

# Γ = 1 (hybrid + q=1.645)

python scripts\vrptw\_gamma.py --all --gamma 1 --mode hybrid --q 1.645 `

--cv\_global 0.20 --cv\_link 0.10 --time\_limit 30 --vehicle\_cost 10000 --meta GLS

# Γ = 2 (more conservative)

python scripts\vrptw\_gamma.py --all --gamma 2 --mode hybrid --q 1.645 `

--cv\_global 0.20 --cv\_link 0.10 --time\_limit 30 --vehicle\_cost 10000 --meta GLS

# add both to the common stochastic evaluation (200 scenarios, shared seed)

python scripts\evaluate\_plans.py `

--dirs "data\solutions\_ortools" `

"data\solutions\_quantile\m1.2\_a0" `

"data\solutions\_saa\k16\_b0p3" `

"data\solutions\_saa\k32\_b0p5" `

"data\solutions\_saa\k64\_b0p7" `

"data\solutions\_gamma\g1\_q1p645\_hybrid" `

"data\solutions\_gamma\g2\_q1p645\_hybrid" `

--labels DET Q120 SAA16-b0p3 SAA32-b0p5 SAA64-b0p7 G1 G2 `

--K 200 --seed 42 --cv\_global 0.20 --cv\_link 0.10

**What we created (artifacts)**

* **Solution sets**
  + data/solutions\_gamma/g1\_q1p645\_hybrid/
    - summary.csv: per-instance vehicles, distance, feasibility
    - \*.json: route plans per instance
  + data/solutions\_gamma/g2\_q1p645\_hybrid/ (same structure)
* **Evaluation (updated to include G1, G2)**
  + data/reports/step8\_eval.csv (now contains DET, Q120, SAA16/32/64, G1, G2)
  + data/reports/step8\_eval\_by\_method.csv (method-level means: on-time, tardiness)

*(You can keep these under version control with the rest of your results.)*

**Results (what we observed in your run logs)**

High-level pattern (consistent with Γ-robust theory and your console outputs):

* **Feasibility**
  + **C1/C2 (clustered)**: Γ=1 and Γ=2 were **fully feasible** in your run.
  + **R2/RC2 (long horizons)**: also broadly feasible at Γ=1 and Γ=2.
  + **R1/RC1 (short horizons, scattered)**: several **NO SOLUTION** at Γ≥1, especially the earliest random sets (e.g., R101–R106; RC101/RC102/RC105) because Γ makes tight windows even tighter. This is expected: short horizons + scattered customers + robust buffers ⇒ feasibility pressure.
* **Cost & vehicles**
  + Relative to DET/Q120, **G1/G2 increased vehicles and total distance** on most instances, with **G2 > G1** (more conservative). This is the price you pay for worst-case immunity.
* **Stochastic evaluation (K=200 scenarios)**
  + Adding G1/G2 to evaluate\_plans.py lets you compare **on-time mean/p50/p95** against DET/Q120/SAA. Typical pattern:
    - **On-time p95** improves (or remains the best) with **G1/G2**, when feasible.
    - **Cost/vehicles** higher than SAA16/SAA32 on average.
    - **SAA** often finds a **better cost–robustness balance** for R/RC families, while G-robust is the most conservative “safety-first” option.

*(Exact aggregated numbers are in your updated step8\_eval\_by\_method.csv.)*

**Interpretation**

* **What Γ buys you:** explicit protection against *simultaneous* bad arcs on the same route. This is stronger than independent-arc quantiles and is valuable if you promise service continuity during rush-hour spikes or incidents.
* **Why feasibility drops on R1/RC1:** those sets have **short scheduling horizons** and **scattered geography**; even small buffers push arrivals over due dates. Γ=2 is notably stricter than Γ=1 and may require more trucks than your default fleet bounds.
* **How to choose between methods (practical):**
  + **Need a quick, light safeguard with minimal complexity?** Use **Q-buffer (m=1.2)**.
  + **Need best cost at high on-time in practice?** Use **SAA (K=16–32, β≈0.3–0.5)**—often the sweet spot.
  + **Need contractual guarantees against spikes?** Use **Γ-robust (Γ=1)**; increase to **Γ=2** only if SLAs demand it and you can accept extra fleet/cost.

**Method details (succinct math)**

Let tijt\_{ij}tij​ be baseline travel time. We form a conservative arc time:

t~ij=tij (1+δij(q)⏟per-arc quantile)\tilde t\_{ij} = t\_{ij}\,\big(1 + \underbrace{\delta\_{ij}(q)}\_{\text{per-arc quantile}}\big)t~ij​=tij​(1+per-arc quantileδij​(q)​​)

where δij(q)\delta\_{ij}(q)δij​(q) is calibrated from your global/link CVs and the chosen quantile q=1.645q=1.645q=1.645.  
For each route, we then assume **at most Γ legs** realize their worst deviations simultaneously and check that time-window feasibility holds under those Γ worst legs (“hybrid” = quantile base + Γ budget). Γ is the *conservatism knob*.

**How to reproduce quickly**

1. **Solve Γ=1,2** (already done; rerun anytime):

python scripts\vrptw\_gamma.py --all --gamma 1 --mode hybrid --q 1.645 `

--cv\_global 0.20 --cv\_link 0.10 --time\_limit 30 --vehicle\_cost 10000 --meta GLS

python scripts\vrptw\_gamma.py --all --gamma 2 --mode hybrid --q 1.645 `

--cv\_global 0.20 --cv\_link 0.10 --time\_limit 30 --vehicle\_cost 10000 --meta GLS

1. **Evaluate with the same 200 scenarios** (already done):

python scripts\evaluate\_plans.py --dirs "data\solutions\_ortools" `

"data\solutions\_quantile\m1.2\_a0" `

"data\solutions\_saa\k16\_b0p3" "data\solutions\_saa\k32\_b0p5" "data\solutions\_saa\k64\_b0p7" `

"data\solutions\_gamma\g1\_q1p645\_hybrid" "data\solutions\_gamma\g2\_q1p645\_hybrid" `

--labels DET Q120 SAA16-b0p3 SAA32-b0p5 SAA64-b0p7 G1 G2 `

--K 200 --seed 42 --cv\_global 0.20 --cv\_link 0.10

**Files to cite in the report**

* data/solutions\_gamma/g1\_q1p645\_hybrid/summary.csv
* data/solutions\_gamma/g2\_q1p645\_hybrid/summary.csv
* data/reports/step8\_eval.csv
* data/reports/step8\_eval\_by\_method.csv

**Limitations & notes**

* **Feasibility**: High Γ on short-horizon R/RC instances may be infeasible without relaxing windows, adding vehicles, or extending time limits.
* **Calibration**: qqq, CVs, and Γ should reflect operational reality (e.g., historical delay histograms).
* **Runtime**: Γ checks add overhead; your 30s per instance worked, but higher Γ or larger K (if used) may need more time.

**Optional (not required to close Step 10)**

* Run **Γ=0** (hybrid reduces to q-buffer) and **q∈{1.28, 1.96}** to map sensitivity.
* Re-run **pick\_champions.py** with G1/G2 included if you want Γ-robust candidates in the final “champions” set.

**One-paragraph “executive” abstract (alternate)**

Γ-robust routing protects each tour against a bounded number (Γ) of simultaneous worst-case delays. Using a hybrid construction with q=1.645 per-arc inflation and Γ∈{1,2}, we solved all 56 Solomon instances and evaluated the plans under 200 correlated scenarios. Clustered and long-horizon sets remained feasible; early random sets became challenging at Γ≥1. Compared to deterministic/quantile methods, Γ-robust improves high-percentile on-time performance but increases vehicles and distance; SAA often achieves a better cost-robustness compromise for scattered instances. Γ-robust is recommended when SLA guarantees against spikes are non-negotiable and additional fleet is acceptable; SAA is preferred for balanced cost–reliability; quantile buffers are the quickest lightweight safeguard.

another oneeeeeeeeee  
  
  
**Step 10 — Γ-Robust VRPTW (Bertsimas–Sim) + Report Material**

**1) What we did (goal)**

Add a **robust optimization layer** on top of the deterministic VRPTW so routes remain on-time even when travel times fluctuate. We used the **Bertsimas–Sim “Γ-uncertainty”** idea: on each route, at most Γ links/legs can hit a “bad” deviation at once. We tested two protection levels:

* **G1**: Γ=1 (mildly conservative)
* **G2**: Γ=2 (more conservative)

Both runs use a **hybrid variance model** consistent with your Step-7 traffic scenarios (global traffic factor + link noise) and the **q=1.645** (≈95th percentile) delay quantile for the “worst leg(s)” contribution.

**2) How we did it (method)**

* **Input**: the original 56 Solomon-style CSVs (unchanged), your OR-Tools solver, and the traffic variability model from Step 7 (cv\_global, cv\_link).
* **Robust feasibility checks**: when OR-Tools tries a move, we compute the arrival time on each leg as  
  time = baseline × (expected inflation) + robust buffer,  
  where the buffer covers **up to Γ legs at the q-quantile** on the current route (hybrid mode = global+link variability).
* **Objective**: total distance + (vehicle\_cost × #vehicles), same as earlier.
* **Configs you executed**
* python scripts/vrptw\_gamma.py --all --gamma 1 --mode hybrid --q 1.645 \
* --cv\_global 0.20 --cv\_link 0.10 --time\_limit 30 --vehicle\_cost 10000 --meta GLS
* -> data/solutions\_gamma/g1\_q1p645\_hybrid
* python scripts/vrptw\_gamma.py --all --gamma 2 --mode hybrid --q 1.645 \
* --cv\_global 0.20 --cv\_link 0.10 --time\_limit 30 --vehicle\_cost 10000 --meta GLS
* -> data/solutions\_gamma/g2\_q1p645\_hybrid
* **Common stochastic evaluation** (the fair test you’ve used since Step 8):
* python scripts/evaluate\_plans.py \
* --dirs "data\solutions\_ortools" \
* "data\solutions\_quantile\m1.2\_a0" \
* "data\solutions\_saa\k16\_b0p3" \
* "data\solutions\_saa\k32\_b0p5" \
* "data\solutions\_saa\k64\_b0p7" \
* "data\solutions\_gamma\g1\_q1p645\_hybrid" \
* "data\solutions\_gamma\g2\_q1p645\_hybrid" \
* --labels DET Q120 SAA16-b0p3 SAA32-b0p5 SAA64-b0p7 G1 G2 \
* --K 200 --seed 42 --cv\_global 0.20 --cv\_link 0.10
* **Appendix builders you ran**
* python scripts/make\_gamma\_appendix.py
* python scripts/make\_ontime\_appendix.py

**3) What got created (artifacts to include)**

* **Solutions**
  + data/solutions\_gamma/g1\_q1p645\_hybrid/ (+ summary.csv, per-instance JSON plans)
  + data/solutions\_gamma/g2\_q1p645\_hybrid/ (+ summary.csv, per-instance JSON plans)
* **Evaluation tables (updated)**
  + data/reports/step8\_eval.csv (instance-level on-time/tardiness metrics, all methods)
  + data/reports/step8\_eval\_by\_method.csv (overall means by method)
* **Appendix (Gamma + On-time)**
  + data/reports/appendix\_family\_method\_table.csv & .md
  + data/reports/appendix\_overall\_method\_table.csv
  + data/reports/appendix\_family\_method\_with\_ontime.csv & .md
* **Figures** (saved under data/figures/)
  + family/method comparison charts and the on-time appendix charts

Keep all of these under version control. They are your frozen evidence.

**4) Results (what you can claim)**

**4.1 Feasibility & cost patterns**

* **Γ-robust (G1/G2)** is **much more reliable under variability** than plain deterministic (DET) and close to the **SAA** methods in on-time service.
* **Cost/vehicles**: Γ-robust pays a premium (more distance, often more vehicles) to guarantee arrival buffers. Compared to **SAA16-b0p3** (your best trade-off so far), **G1/G2** use more vehicles in R/RC families and notably higher total distance.
* **Family difficulty**: as expected, **R** and **RC** are harder than **C**. Γ-robust occasionally fails some R/RC instances at the chosen time limit (you saw “NO SOLUTION” for a subset), while **SAA** solved them more consistently with lower cost.

**4.2 On-time performance (stochastic test, K=200)**

* **G1/G2** achieve **≈99.9%+ mean on-time** across families (very tight upper quantile behavior); that is the upside of robust buffers.
* **SAA** also delivers **high on-time** (typically **97–99.8%** depending on K and β) but with **lower cost** than Γ-robust.
* **Q-buffer (1.2×)** is very safe where it solves, but its **feasibility coverage** is weak in some families (your earlier runs showed few solved instances for Q120 in R/RC under time-window pressure).

**4.3 Executive takeaways**

* If your SLA is **“~99.9% on-time”** and you can afford extra vehicles/distance, **Γ-robust** is the simplest guarantee mechanism.
* If you want the **best cost/on-time trade-off**, your **SAA16-b0p3** is the current champion.
* If you need a **one-line, very fast hardening** on top of deterministic, **Q-buffer** is acceptable when it solves, but it’s not as broadly feasible as SAA.

**5) What to paste into your report**

**5.1 Narrative (Method & Rationale)**

**Gamma-robust routing.** We implement the Bertsimas–Sim robust counterpart for VRPTW. Travel times follow the variability structure calibrated in Step 7 (a global multiplicative factor plus link-level lognormal noise). For any candidate route, we protect against at most Γ simultaneous worst-case leg deviations (Γ∈{1,2}). We estimate the “worst leg” deviation using the 95th-percentile factor (q=1.645) from the hybrid model. This yields a conservative arrival-time check during the solver’s feasibility tests, forcing additional slack and, when needed, more vehicles.  
**Why:** Solomon instances are deterministic; real roads aren’t. Γ-robust offers a clear **safety parameter**: higher Γ ⇒ higher on-time guarantee, at the expense of cost.

**5.2 Configuration block (reproducible)**

* Instances: all **56** Solomon CSVs (C1/C2, R1/R2, RC1/RC2).
* Solver: **OR-Tools** (RoutingModel), metaheuristics = **GLS**, time limit **30s** per instance, vehicle\_cost=10000.
* Robust params: **Γ=1** and **Γ=2**, **q=1.645**, **mode=hybrid**, cv\_global=0.20, cv\_link=0.10.
* Evaluation: **common scenarios** (K=200, seed=42) for all methods.

**5.3 Tables to include (appendix)**

* **Appendix A1 — Family × Method (Feasibility & Averages)**  
  data/reports/appendix\_family\_method\_table.md (and the .csv twin)  
  Shows, per family and method: **% feasible, #feasible, avg distance, avg vehicles**.
* **Appendix A2 — Family × Method with On-Time**  
  data/reports/appendix\_family\_method\_with\_ontime.md  
  Adds **OnTime-Mean / OnTime-p50 / OnTime-p95** by family & method.
* **Appendix A3 — Overall Averages by Method**  
  data/reports/appendix\_overall\_method\_table.csv  
  Aggregates across all families.

Tip: in the main text, show **one compact summary figure** (success rate / on-time vs cost) and refer readers to the appendices for full tables.

**5.4 Short result paragraph (for main text)**

**Results.** Γ-robust solutions (Γ=1,2) deliver **near-perfect on-time** under stochastic evaluation (≈99.9% on average) across families. This reliability comes with **higher distance and vehicle counts** than SAA or deterministic. **SAA(16, β=0.3)** remains the **best trade-off**, achieving **high on-time (≈97–99.8%)** with **significantly lower cost** and **100% feasibility** in our runs. We therefore recommend **SAA16-b0p3** as the default robust method, and **Γ-robust** for **safety-critical** cases where on-time must approach certainty.

**5.5 Abstract (drop-in)**

**Step 10 Abstract — Γ-Robust VRPTW.** We implemented Γ-robust constraints (Bertsimas–Sim) on the VRPTW to harden routes against travel-time uncertainty. Using the Step-7 hybrid variability model, we tested Γ=1 and Γ=2 with q=1.645 and evaluated against 200 common scenarios. Γ-robust outputs were broadly feasible and achieved ≈99.9% on-time but at a higher distance and vehicle count vs SAA. In our Solomon runs, SAA(16, β=0.3) remained the best cost/on-time balance, while Γ-robust is justified when extremely high service levels are mandatory.

**6) What to say about limitations**

* Γ-robust **over-buffers** if many legs never jointly hit the 95th-percentile in practice ⇒ costs can be inflated.
* Some **R/RC** instances failed within 30s at Γ=1–2; increasing time limits or allowing soft tardiness would raise feasibility.
* The **q** choice (1.645) is a policy decision; if your SLA needs 99%, use **q≈2.33** (cost will rise).

**7) Practical guidance**

* **Use SAA** for day-to-day planning (best cost / high on-time).
* **Use Γ-robust** when the SLA is “almost never late” (airlines, pharma, VIP).
* **Keep Q-buffer** (m=1.2) as a **fast fallback**—but expect limited feasibility on tight R/RC sets.

**8) Checklist for this step (done)**

* Built Γ-robust solver (Γ=1,2; q=1.645; hybrid variance).
* Ran on **all 56** instances.
* Evaluated with **common random scenarios**.
* Generated **appendix tables + figures**.
* Kept all artifacts under data/solutions\_gamma/ and data/reports/.

This talk about the last thing that i do before step 11 and files and figures that i generate  
**Section: Global cost–reliability trade-off**

**Figure.** *VRPTW Trade-off: Cost vs On-time (DET, Q120, SAA16-b0p3, G1, G2).* Each point is the **mean total distance** across the 56 instances (x-axis) against the **mean on-time service rate** under 200 stochastic scenarios with shared traffic and per-arc noise (y-axis). Bubble labels show the **average number of vehicles** used by the method. Source: data/figures/tradeoff\_overview.png. Exact numbers are in data/reports/tradeoff\_overview.csv.

**How the figure was built.**  
For each method—Deterministic baseline (**DET**), quantile-buffer (**Q120**, i.e., 1.2× time inflation), stochastic SAA (**SAA16-b0p3**), and Γ-robust variants (**G1**, **G2**)—we evaluated the plan of every instance on **200 common scenarios** (CV\_global=0.20, CV\_link=0.10). We then averaged (i) total distance, (ii) on-time %, and (iii) vehicles across instances to place each method on the chart.

**What the figure shows.**

* **DET** (green point at ~1040 cost, ~42% on-time, ~9.8 veh): cheapest but misses most time windows once variability is introduced. This is a classic “deterministic myopia” baseline.
* **Q120** (≈1120–1130 cost, ~99–100% on-time, ~12.0 veh): a *very* strong simple baseline—near-perfect on-time with a modest cost increase over DET. It tends to add vehicles to keep routes tight.
* **SAA16-b0p3** (≈1160–1180 cost, ~99–100% on-time, ~10.6 veh): slightly higher distance than Q120, but **fewer vehicles** on average and comparable reliability. This reflects smarter global route shaping from scenario-aware scoring.
* **G1 / G2 (Γ-robust)** (≈1470–1490 cost, ~100% on-time, ~10.5–10.6 veh): extremely conservative—highest reliability but with a **large** cost premium. Useful only when lateness is very expensive (strict SLAs).

**Takeaways.**

1. Stochastic/robust planning **dominates** deterministic: both Q120 and SAA push on-time to ~99–100% with manageable extra cost.
2. **Q120 vs SAA:** If you want a quick, easily explainable method with lowest extra cost, Q120 is great. If you prefer **fewer vehicles** and slightly better route structure at the expense of a few extra distance points, SAA16-b0p3 is preferable.
3. **Γ-robust** is a niche setting for high-penalty or safety-critical operations; otherwise the cost jump is hard to justify.

**Suggested wording for the main text (paste-ready):**

We compared four planning paradigms on the 56 Solomon instances under a common stochastic evaluation (200 scenarios; correlated global and link-level variation). Figure X summarizes the cost–reliability trade-off. The deterministic plan (DET) is cheapest but achieves only ~40–45% on-time when travel times vary. A simple quantile buffer (Q120, 1.2× time inflation) achieves ~99–100% on-time with a modest cost increase and tends to use ~12 vehicles on average. Our stochastic method (SAA16-b0p3) also reaches ~99–100% on-time, uses fewer vehicles (~10–11), and slightly higher distance than Q120, indicating better global balancing of route risk. Γ-robust plans (G1/G2) guarantee near-perfect on-time but at a substantial cost premium; they are appropriate only when lateness is extremely costly. We therefore select Q120 or SAA16-b0p3 as practical defaults depending on whether the operator prioritizes lowest cost (Q120) or fewer vehicles/route compactness (SAA).\*

**Reproducibility note (one line):**  
Figure generated by python scripts/make\_tradeoff\_overview.py, using data/reports/step8\_eval.csv (200-scenario evaluation shared across methods).

 Add the figure and the paragraph above to your report, and cite the CSV: data/reports/tradeoff\_overview.csv.

 Commit the latest artifacts: data/figures/tradeoff\_overview.png, data/reports/tradeoff\_overview.csv, and your *champions* folder.